An Efficient, Differentiable Hydration Potential for Peptides and Proteins

JOSEPH D. AUGSPURGER* and HAROLD A. SCHERAGA†
Cornell University, Baker Laboratory of Chemistry, Ithaca, New York 14853-1301

Received 12 September 1995; accepted 20 November 1995

ABSTRACT

An approximate method for calculating the exposed volume of the hydration shell (VHS) about an atom, the **R**educed **R**adius **I**ndependent **G**aussian **S**phere (RRIGS) approximation, is presented. A key ingredient in this method is the use of reduced van der Waals radii so that the error of including only double overlap terms (and omitting multiple overlap terms) in calculating the VHS is balanced by a reduction in the magnitude of the double overlap terms. Also, the double overlap is modeled with a gaussian function. The RRIGS approximate calculation of the VHS is shown to be very accurate (the rms deviation of the VHS of each atom in avian pancreatic polypeptide and bovine pancreatic trypsin inhibitor was 14.0 and 15.8 $Å^3$, respectively, out of a range of values between 0 and 600 $Å^3$). The RRIGS approximation is used to develop a potential function to represent the free energy of solvation for proteins. The pairwise gaussian form of the potential enables it to be incorporated into a gaussian representation of ECEPP (Empirical Conformational Energy Program for Peptides) for use in the Diffusion Equation Method (DEM) of global optimization. Inclusion of the effects of hydration by means of this potential is shown to require less than twice the computational time needed for computing the ECEPP conformational potential energy alone; this makes inclusion of solvent computationally feasible. Furthermore, this gaussian hydration potential function and its derivatives are continuous, so that it may be readily minimized. The combined potential of ECEPP/3 plus hydration is shown to require fewer energy evaluations per local minimization than ECEPP/3 alone for two small peptides.

Introduction

It is widely believed that the three-dimensional structure of a polypeptide or protein is the thermodynamically most stable conformation.[1,2] Based on this assumption, the prediction of the structure of a protein can be accomplished by minimizing its conformational energy function. Successful application of this approach requires a function which represents the intra- and intermolecular interactions accurately and the means to find the conformation of lowest energy. The diffusion equation method (DEM) has shown great promise for overcoming the multiple-minima

*Special Fellow of the Leukemia Society of America.
†Author to whom all correspondence should be addressed.

CCC 0192-8651 / 96 / 131549-10

problem in global optimization for peptides.[3] Essentially, this method deforms the function, $f(\mathbf{x})$, to be minimized by treating it as the initial condition of a diffusion problem. A solution to the diffusion equation

$$\Delta F(\mathbf{x}, t) = \frac{\partial}{\partial t} F(\mathbf{x}, t) \tag{1}$$

with the initial condition $F(\mathbf{x}, 0) = f(\mathbf{x})$, can be obtained by evaluating the Fourier–Poisson integral,

$$F(\mathbf{x}, t) = \frac{1}{(\sqrt{4\pi t})^d} \int f(\mathbf{y}) \exp\left[-\frac{1}{4t} \|\mathbf{x} - \mathbf{y}\|^2 \right] d\mathbf{y} \tag{2}$$

where d is the number of components in the vector $\mathbf{x}$, i.e., the number of variables. The integral in eq. (2) can be evaluated analytically for certain forms of $f(\mathbf{x})$, e.g., for a sum of gaussians. The deformed function is identified with $F(\mathbf{x}, t)$, where the extent of deformation is associated with time (t). For large deformation times, only a few minima will remain; these local minima can be found easily by other global optimization methods. Then, a time-reversal procedure is applied to follow these few minima of the deformed function back to corresponding local minima of the original function. The global minimum conformation is likely to be among the local minima in this set.[3,4]

The DEM has been applied to the five-residue peptide, met-enkephalin,[3] using a gaussian adaptation of the ECEPP energy function.[5–7] However, since no method (compatible with the DEM) was available for including hydration, the interaction of the peptide with water was omitted in that initial application.[3] The goal of the present work is to develop a representation of the free energy of hydration which can be treated by the DEM.

It would be ideal to incorporate highly accurate potentials which go beyond just pairwise interactions[8] while carrying out simulations with explicit water molecules. However, such an approach is computationally unfeasible for folding a protein. Thus, many different methods have been developed for representing the free energy of hydration of proteins. These methods make the general assumptions that the total hydration free energy is a sum of atomic contributions, and that these atomic contributions are proportional to the solvent exposure of the individual atoms. Gibson and Scheraga introduced the hydration shell model,[9] in which the hydration free energy of a given atom was based on the number of solvent (water) molecules excluded from the hydration shell about the given atom by the other atoms in the protein. This model was modified by Hopfinger,[10] and extended to include specific protein–solvent hydrogen bonds.[11] Further work by Kang et al.[12,13] represented the free energy of a given atom as proportional to the solvent *exposed* volume of the hydration shell, and made use of a more exact calculation of the hydration shell volume which was made possible by the derivation of analytical formulae for the volume of intersection of three and four spheres.[12,14]

Another approach is based on the solvent accessible surface area (SASA) as a measure of solvent exposure. Many numerical[15–20] and analytical[21–31] methods have been presented for calculating the SASA and developing hydration potentials.[32–37] The numerical approach has been shown to be quite rapid in terms of computational time; however, it is unsuited for the calculation of the gradient of the SASA, which is necessary for molecular dynamics simulation or accurate minimization algorithms. Formulae for these gradients based on analytical approaches have been presented,[22,27] and two approaches[27,28] have been shown to rival the numerical approaches in speed. However, the gradient of the SASA suffers from inherent discontinuities,[38] which lead to numerical difficulties in actual practice. The recent report of Sridharan et al.[31] describes a numerical approximation to the analytical approach to surmount this problem. Unfortunately, none of these approaches can be transformed analytically by the Fourier–Poisson integral.

We report here a method for representing the hydration interaction which meets the mathematical requirements of the DEM [analytical solution of eq. (2)] and also avoids the discontinuities of the gradient which cause difficulties in minimization or molecular dynamics simulation. The resulting potential is also shown to be quite computationally efficient. The most recent experimental free energy of hydration data are used in the empirical parameterization of the model.

Theory and Methods

HYDRATION SHELL MODEL

The approach taken here to describe the interaction of the protein with water is the hydration shell approximation.[9–13] Following the application of this model by Kang et al.,[12,13] the total interac-

tion free energy is represented as a sum of free energies of the interaction of each individual atom with water, and for each atom, the hydration interaction free energy is proportional to its exposure to water. Thus:

$$\Delta G_{\text{hyd}} = \sum_i \delta_i (\text{VHS})_i \tag{3}$$

where δ_i is an empirically determined free energy of hydration density for atom i, and $(\text{VHS})_i$ is the volume of the hydration shell about atom i which is exposed to water. The hydration shell is defined as the volume inside a sphere of radius R_i^h and outside a sphere of radius R_i^v centered on atom i. R_i^h is the radius of the first hydration shell of atom i, and R_i^v is its van der Waals radius (see Fig. 1). The protein itself is represented by a set of overlapping hard spheres located at the atomic centers. The radius of each sphere is the van der Waals radius of the particular atom. The volume of such a collection of overlapping hard spheres can be written as:

$$V = \sum_i a_i S_i - \sum_{ij} b_{ij} D_{ij} + \sum_{ijk} c_{ijk} T_{ijk} - \sum_{ijkl} d_{ijkl} Q_{ijkl} \tag{4}$$

where S represents the volume of a single sphere and D, T, and Q the volume of intersection of two, three, and four spheres, respectively. This expression is exact because all higher order overlaps can be written in terms of fourth and lower order overlaps (see Ref. 14, and refs. therein); thus, the determination of the coefficients a, b, c, and d requires examination of all possible higher order overlaps. The volume of the exposed hydration shell of atom i can then be written as:

$$(\text{VHS})_i = V(R_i^h) - V(R_i^v) \tag{5}$$

where $V(R_i^v)$ is the volume of the protein [computed by eq. (4)] when the radii of all atoms are set to their van der Waals radii; $V(R_i^h)$ is the volume [computed by eq. (4)] when the radius of atom i is set to R_i^h and the radii of all other atoms are set to their van der Waals radii.

This approach [based on eq. (4)] is unsuited to transformation by means of the Fourier–Poisson integral; however, such a transformation would be possible if the function V could be reduced to a sum of independent pairwise terms. Such a direct truncation of eq. (4) at the double-overlap term would lead to large errors. One way to alleviate this problem has been suggested by Hodes et al.,[11] namely, reduce the van der Waals radii of all atoms *other* than atom i artificially when calculating $(\text{VHS})_i$. This reduction of all the *other* van der Waals radii decreases the double overlap terms, leading to a cancellation of the error which results from neglecting the triple and higher overlap terms (see Fig. 2). Thus, the *exposed* volume of the hydration shell about atom i can be approximated by:

$$(\text{VHS})_i \approx \frac{4\pi}{3}\left(R_i^{h^3} - R_i^{v^3}\right) - \sum_{j \neq i}\left[D\left(r_{ij}; R_i^h, R_j^r\right) - D(r_{ij}; R_i^v, R_j^r)\right] \tag{6}$$

where $D(r_{ij}; R_i, R_j)$ is the volume of intersection of two spheres of radius R_i and R_j whose centers are separated by r_{ij}, and R_j^r is a reduced van der Waals radius for atom j. This reduction of van der Waals radii is carried out only for the calculation of the hydration free energy; for the nonbonded (Lennard–Jones) part of ECEPP, the original van der Waals radii are used. The use of reduced radii has also been applied to the development of a pairwise calculation of the electrostatic component of the hydration interaction by means of the generalized Born approximation.[39]

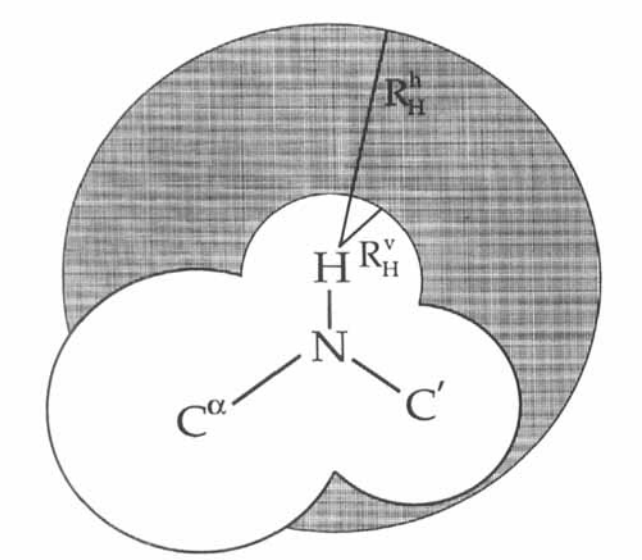

FIGURE 1. The shaded area is the exposed volume of the hydration shell about an amide hydrogen. The hydration shell is bounded by the spheres about the amide hydrogen of radius R_H^h and R_H^v, respectively, and the exposed volume of the hydration shell (VHS) is that volume not occupied by other atoms of the protein (here, the C^α, N, and C′ atoms).

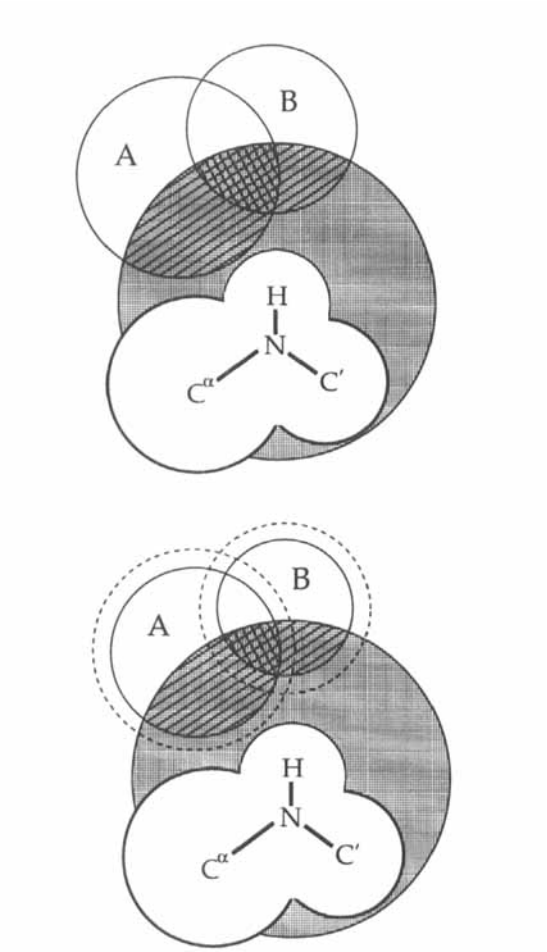

FIGURE 2. Illustration of construction with reduced radii to compensate for the error introduced by truncating the volume calculation of eq. (4) after the double overlap terms. In the upper diagram, the volume of the hydration shell of the H atom, $(VHS)_H$, occluded by atoms A and B [i.e., $\Sigma_{j \neq i}[D(r_{ij}; R_i^h, R_j^r) - D(r_{ij}; R_i^v, R_j^r)]$ of eq. (6)] is indicated by upward sloping lines (///). This volume can be calculated exactly as $D_{H(hyd),A} + D_{H(hyd),B} - T_{H(hyd),A,B}$ [the sum of the double overlaps minus the triple overlap (the latter being indicated by cross-hatching)]. The lower diagram shows the same conformation, except that the van der Waals radii of *A* and *B* have been reduced (with the original volume drawn with dashed lines). Here, the sum of the double overlaps of the hydration shell of the H atom with the reduced van der Waals atoms *A* and *B* is exactly the same as the volume calculated above. This arises because the triple overlap term in the lower diagram (indicated by cross-hatching) *exactly* equals the reduction in the double overlap terms. It should be noted that, in this example, there is no overlap between the van der Waals spheres of the atoms which are not bonded to *A* and *B*, i.e., between *A*, *B*, and H, C^{α}, N, or C'. Because of the repulsive term in the nonbonded potential in ECEPP, such overlaps will be predominantly small. This is crucial to the success of this approximation.

While this approximation will be shown to be appropriate for the hydration shell volume approach, it is not appropriate for calculating the SASA accurately. The SASA of a protein is defined as the locus of points traced out by the center of a probe sphere (representing a water molecule) as it rolls over the van der Waals surface of the protein; this set of points can be defined equivalently as the surface of the protein when each atom has its radius set to the van der Waals radius plus the radius of the probe sphere ($R_i^{SASA} = R_i^v + R^{probe}$). In proteins, the atoms are densely packed, but the nonbonded repulsive potential between atoms prevents them from approaching each other much closer than the sum of their van der Waals radii. However, as a result, the atomic spheres (whose radii are equal to R_i^{SASA}) are larger and interpenetrate significantly.

The area of a single sphere in a collection of spheres is the derivative of the total volume with respect to its radius.[14] Therefore, $(SASA)_i = \partial V / \partial R_i^{SASA}$ [where V, defined by eq. (4), is the volume enclosed by the SASA]. In calculating V [and, hence, $(SASA)_i$, which is derived from V], the triple and quadruple overlap terms may be large. It is unlikely that, by simply reducing R_i^{SASA} and thereby reducing the double overlap terms in eq. (4), such a procedure will accurately compensate for neglecting the triple and quadruple overlap terms.

On the other hand, in the hydration shell volume model, the determination of $(VHS)_i$ involves the calculation of the overlap of all other atomic *van der Waals* spheres with the hydration shell of atom *i*. Since the van der Waals spheres of the other atoms interpenetrate only slightly, neglecting the triple, quadruple, and higher overlap terms [as in eq. (6)] is appropriate. For covalently bonded atoms, whose van der Waals spheres do interpenetrate, the reduced radii approach compensates for the existence of higher order overlaps (see Fig. 2).

GAUSSIAN APPROXIMATION FOR THE VOLUME OF DOUBLE OVERLAPS

In light of the goal of developing a hydration potential which can be transformed by the Fourier–Poisson integral, we introduce a gaussian approximation to $D(r_{ij}; R_i, R_j)$. The volume of a set of intersecting spheres can be calculated by integrating a characteristic density function over all space. For hard spheres, this characteristic density function has the value of 1 inside any sphere, and is zero otherwise. Let ρ_i be this characteristic

density function for sphere i. Then, for two spheres, the composite density function $\mathbf{P}_{12}$ can be written as:

$$\mathbf{P}_{12} = 1 - (1 - \rho_1)(1 - \rho_2) = \rho_1 + \rho_2 - \rho_1 \rho_2 \quad (7)$$

The function $\mathbf{P}_{12}$ possesses the same property for the union of the two spheres as ρ_i possesses for a single sphere; it is 1 inside either sphere, and 0 otherwise. Thus, integrating $\mathbf{P}_{12}$ over all space gives the volume of the united spheres. The volume of intersection of two spheres with radius R_1 and R_2 separated by a distance r_{12} is found by integrating over only the $\rho_1 \rho_2$ term, which leads to:

$$D(r_{12}; R_1, R_2) = \frac{2\pi}{3}\left(R_1^3 + R_2^3 + \frac{r_{12}^3}{8}\right) - \frac{\pi r_{12}}{2}(R_1^2 + R_2^2) - \frac{\pi}{4r_{12}}(R_1^2 - R_2^2)^2 \quad (8)$$

for $|R_1 - R_2| \leq r_{12} \leq (R_1 + R_2)$. Alternatively, we define a different characteristic density function which will represent a "softer" sphere. Instead of a step function ($\rho = 1$ inside the sphere, $\rho = 0$ outside), we choose to represent this soft-sphere density by a gaussian function:

$$\rho(r) = A \exp\left(\frac{-r^2}{BR^2}\right) \quad (9)$$

where r is the distance from the center of sphere, R is its analogous hard-sphere radius, and A and B are empirical scaling parameters to be determined later. Integration as before, now with eq. (9) substituted into the $\rho_1 \rho_2$ term of eq. (7), determines the formula for the volume of intersection of two "gaussian spheres":

$$D_G(r_{12}; R_1, R_2) = A^2\left(\frac{\pi B}{R_1^2 + R_2^2}\right)^{3/2} R_1^3 R_2^3 \exp\left(\frac{-r_{12}^2}{B(R_1^2 + R_2^2)}\right) \quad (10)$$

where, as in eq. (8), R_1 and R_2 are the hard-sphere radii and r_{12} the separation between the centers of the spheres. This expression for D_G can be substituted into eq. (6) in place of the expressions for D (the volume of intersection of two hard spheres) to calculate $(\mathrm{VHS})_i$ approximately. A and B are treated as empirical parameters that can be varied to minimize the difference between this approximate calculation of $(\mathrm{VHS})_i$ and the exact calculation by means of eqs. (4) and (5). We designate this approximation as the **R**educed **R**adius, **I**ndependent **G**aussian **S**phere approximation (RRIGS). Other investigators have also utilized gaussian functions to measure molecular volume and surface area,[41] or solvent exposure about individual atoms.[42]

The reduced van der Waals radii required for the RRIGS calculation of $(\mathrm{VHS})_i$ are determined for a given atom as follows. First, a fixed boundary is drawn between the atom $j(\neq i)$ and each of its bonded neighbors $k(\neq i)$. This boundary was chosen to be the set of points which satisfy the condition $r_j/r_k = R_j/R_k$, where r_j and r_k are the distances from any point to the centers of spheres j and k, and R_j, R_k are the respective van der Waals radii. This boundary between spheres is itself a sphere (except in the case in which $R_j = R_k$, where it defines a plane), as illustrated in Figure 3, and preferentially partitions the volume to the smaller sphere. This boundary definition has recently been applied to the calculation of atomic volumes in proteins.[43] V_j^r, the individual volume assigned to atom j, is that volume of atom j which is inside all of the boundaries between atom j and its

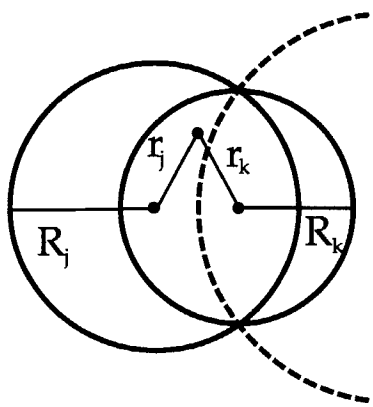

FIGURE 3. Illustration of the partitioning of the volume between overlapping spheres to determine the reduced radii. For two spheres of radii R_j and R_k, the boundary is defined to be the locus of points where $r_j/r_k = R_j/R_k$ (shown by the dashed line), which defines a third sphere. The reduced volume of sphere j (V_j^r) is that volume of atom j **not** contained in the dashed sphere (where, for any point in j, $r_j/r_k < R_j/R_k$) and the reduced volume of sphere k (V_k^r) is that volume of atom k **also contained** in the dashed sphere (where, for any point in k, $r_j/r_k > R_j/R_k$).

covalently bonded atoms. The reduced radius for each atom is then defined to be:

$$R_j^r = (3V_j^r/4\pi)^{1/3} \quad (11)$$

This procedure had to be carried out only once because the ECEPP/3 potential function uses fixed bond lengths and bond angles. Hence, the covalently bonded neighbors of an atom never move with respect to the given atom; therefore, the appropriate reduced van der Waals radius depends on the geometry (bond lengths and bond angles) with respect to the covalently bonded atoms, but is independent of protein conformation. Since the reduced radius depends on the exact geometry of its covalent partners, it must be computed separately, but only once, for all atoms of all amino acids. Therefore, V_j^r was calculated by Monte Carlo integration.

Because of the fixed geometry used in the ECEPP potential function, the contribution of covalently bonded atoms to $(VHS)_i$ is fixed. Thus, this constant contribution to $(VHS)_i$ is precomputed, so that we can finally write

$$(VHS)_i \approx (VHS)_i^{FIX} - \sum_{j \neq \{i\}} \left\{ D_G\left(r_{ij}; R_i^h, R_j^r\right) - D_G(r_{ij}; R_i^v, R_j^r) \right\} \quad (12)$$

where $j \neq \{i\}$ indicates a sum over all atoms j which are neither neighbors nor next-nearest neighbors of atom i. Therefore, the RRIGS approximation has exactly the same set of interactions [viz. (1,4), (1,5), etc.] as in the ECEPP potential, and the hydration free energy can be computed as:

$$\Delta G_{hyd} = \Delta G_{hyd}^{FIX} - \sum_{\{i,j\}} \left[\delta_i \left\{ D_G\left(r_{ij}; R_i^h, R_j^r\right) - D_G(r_{ij}; R_i^v, R_j^r) \right\} + \delta_j \left\{ D_G\left(r_{ij}; R_j^h, R_i^r\right) - D_G(r_{ij}; R_j^v, R_i^r) \right\} \right] \quad (13)$$

where $\{i, j\}$ indicates that the sum is taken over all (1,4), (1,5), and higher-order interacting pairs of atoms (as in ECEPP) and:

$$\Delta G_{hyd}^{FIX} = \sum_i \delta_i (VHS)_i^{FIX} \quad (14)$$

Thus, this hydration potential can readily be incorporated into the existing ECEPP program, and is parallelizable.

COMPARISON OF RRIGS APPROXIMATION TO EXACT VHS

The determination of the parameters A and B in eqs. (9) and (10) was carried out by generating a template peptide consisting of a connected sequence of the 20 naturally occurring amino acids in a fully extended conformation. $(VHS)_i$ was calculated for each atom using the exact expressions of eqs. (4) and (5), and then with the RRIGS approximation. The A and B parameters were determined by a search over the A, B space for the best fit of the approximate values to the exact values. The resulting values are $A = 1.88$ and $B = 0.52$, which are the same for *all* interacting hydration shells. Figure 4 illustrates the results of this fitting, and two test cases using the native conformations of avian pancreatic polypeptide (36 residues) and bovine pancreatic trypsin inhibitor (58 residues). The rms deviation of $(VHS)_i$ for these three cases were 22.4, 14.0, and 15.8 Å^3, respectively. As can be seen from the figures, the values of $(VHS)_i$ for individual atoms range from 0 to 600 Å^3. This agreement for such different conformations (fully extended to fully folded) is an indication of the robustness of the accuracy of this approximate calculation of $(VHS)_i$.

EMPIRICAL PARAMETERIZATION OF MODEL

Both the van der Waals and hydration radii were taken from Kang et al.[12] and are given in Table I. These radii were used in the determination of the values of the reduced radii (R_j^r) by means of eq. (11). The values for the empirical free energy densities, δ_i, in eq. (3) were obtained by a least squares fitting of experimental free energy of solvation data for 140 small organic molecules with eq. (3) using the exact expressions of eqs. (4) and (5). The experimental data were taken from the recent compilation by Abraham et al.[44] The data are given in Table 4 of ref. 44 as $\log(L^W)$, where L^W is the Ostwald solubility coefficient (i.e., concentration of solute in solution/concentration of solute in gas phase), which can be converted to the free energy of hydration by the relation:

$$\Delta G(\text{gas} \rightarrow \text{aqueous}) = -2.303RT \log(L^W) \quad (14)$$

where the standard states are gaseous solute at a concentration of 1 mol/L and solute in an ideal dilute aqueous solution at a concentration of 1 mol/L. Our goal is the development of a potential function to represent the interaction of a protein (in a given conformation) with water. Hence, the

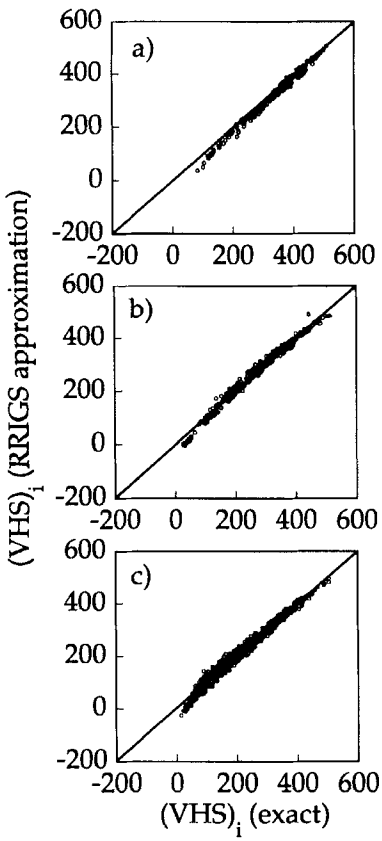

FIGURE 4. A comparison between the exact calculation of (VHS)$_i$ by eqs. (4) and (5) and the approximate RRIGS method by eqs. (5) and (12) for three different peptides: (a) a test peptide consisting of a connected sequence of the 20 naturally occurring amino acids in a fully extended conformation (all $\phi, \psi, \omega = 180°$), which was used to determine the A, B parameters; (b) avian pancreatic polypeptide (36 residues) in its native conformation; and (c) bovine pancreatic trypsin inhibitor (58 residues) in its native conformation. In each diagram, the computed (VHS)$_i$ is represented by an open circle, and the solid line indicates exact agreement between the two methods.

transfer of the solute from *gas* to aqueous solution is the appropriate physical process to model such interactions.[33]

Since the compounds in the compilation of Abraham et al.[44] can exist in several conformational states, an appropriate (Boltzmann weighted) set of conformations must be used to compute (VHS)$_i$. This computation was carried out in several steps to achieve self-consistency. In the initial step, (VHS)$_i$ was calculated for an ensemble of conformations which consisted of all rotamers about single bonds (gauche$^+$, gauche$^-$, trans) and double bonds (cis, trans). While only one cis or trans isomer, but not both, is likely to contribute

TABLE I.
Empirical Free Energy Density of Solvation Parameters.

i[a]	Atom type	$\delta_i \times 10^3$[b]	R_i^v (Å)[c]	R_i^h (Å)[c]
1	hydroxyl, amino **H**	−10.35	1.415	4.17
2	acid **H**	−3.206	1.415	4.17
3	amide **H**	−7.714	1.415	4.17
4	thiol **H**	2.709	1.415	4.17
5	aliphatic **C**H_3	1.319	2.125	5.35
6	aliphatic **C**H_2	0.2374	2.225	5.35
7	aliphatic **C**H	−1.271	2.375	5.35
8	aliphatic **C**	−2.297	2.06	5.35
9	cyclic **C**R	0.3827	2.25	5.35
30	cyclic **C**H	0.2890	2.375	5.35
10	aromatic **C**H	−0.2137	2.10	5.35
11	aromatic **C**R	−1.713	1.85	5.35
12	branched aromatic **C**	−1.910	1.85	5.35
13	aromatic **C**OH	−0.6063	1.85	5.35
14	carbonyl **C**	2.696	1.87	5.35
15	primary amine **N**	−1.149	1.755	5.05
17	secondary amine **N**	−10.28	1.755	5.05
20	aromatic **N**	−10.48	1.755	5.05
22	amide **N**	−7.332	1.755	5.05
23	hydroxyl, ether **O**	−7.396	1.62	4.95
24	acid, ester **O**	0.07897	1.62	4.95
25	ketone carbonyl **O**	−15.70	1.56	4.95
26	acid, amide carbonyl **O**	−15.56	1.56	4.95
28	thiol, disulfide **S**	−4.706	2.075	5.37

[a] This numbering coincides with the numbering of prior hydration potentials[37] designed to be added to ECEPP.
[b] In kcal/mol/Å^3, where the reference states are gaseous solute at a concentration of 1 mol/L and solute in an ideal aqueous solution at a concentration of 1 mol/L.
[c] Taken from Ref. 12.

(except perhaps for amides), they are both included for completeness. This weighting was carried out by using an ECEPP-like energy function[5-7] for the molecules in the database. This energy function has the same functional forms as ECEPP/3. The parameters were generated by calculating partial charges of the atoms for each molecule by means of CNDO calculations (as in ECEPP) and the nonbonded parameters were taken directly from ECEPP/3.

With these Boltzmann-averaged values of (VHS)$_i$ for each molecule, the δ_i values were determined by a linear least squares fitting of the data to eq. (3). These initial values of the δ_i values were then used in the subsequent iteration, where the Boltzmann averaging was now carried out with an energy which consisted of the ECEPP-like

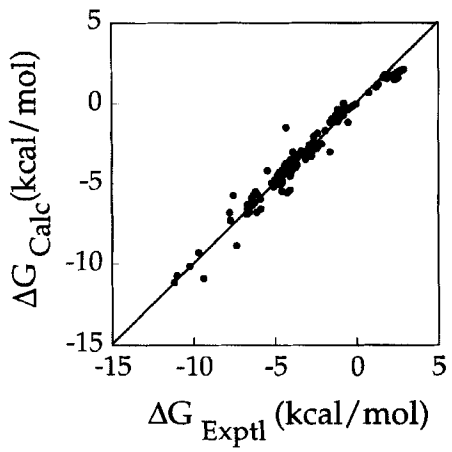

FIGURE 5. Comparison of the experimental values of ΔG(gas → aqueous)[44] from eq. (14) with the values calculated by means of eq. (3) for the 140 small organic molecules used in the fitting process to determine the free energy density parameters (δ_i).

potential plus the free energy of hydration. This procedure was repeated, but self-consistency had already been achieved in the first iteration. The resulting parameters from this self-consistent fitting are given in Table I. The fitted values are compared to the experimental values in Figure 5. The average absolute deviation for the 140 compounds used in determining the empirical parameters was 0.35 kcal/mol and the standard deviation was 0.53 kcal/mol.

Results and Discussion

The RRIGS hydration potential has been added to ECEPP/3 to form a total conformational energy function:

$$E_{TOT} = \text{ECEPP/3} + \Delta G_{hyd}(\text{RRIGS}) \quad (15)$$

which has been implemented in a version of the ECEPP program on an IBM SP2 computer. Timing comparisons have been made between ECEPP/3, E_{TOT}, and another method for including hydration, ECEPP/3 + MSEED, which has been shown to be computationally efficient.[27] Comparisons were made between these potential functions for three different polypeptides consisting of 5, 15, and 58 amino acids each. The results are given in Table II. In these calculations, the time reported includes that for calculating both the energy and the cartesian gradients. For all three peptides, E_{TOT} requires less than twice the time of ECEPP/3 alone. As would be expected, since both potentials have exactly the same number of interactions to calculate, they scale similarly; i.e., the ratio of the number of energy evaluations per second for E_{TOT} to that of ECEPP/3 is independent of the size of the peptide.

Compared to the MSEED algorithm, E_{TOT} is faster for the smaller peptides, but slightly slower for bovine pancreatic trypsin inhibitor (BPTI). This illustrates one strength of the MSEED algorithm—that it deals only with the surface of the protein and so scales less strongly with the size of the protein. Yet, E_{TOT} is quite comparable in speed to this rapid algorithm, does not suffer from the discontinuities of the gradient or the neglect of interior cavities, and is suitable for minimization by means of the DEM algorithm.

Another question of interest is whether the nature of the conformational energy landscape of ECEPP/3 is altered by the inclusion of hydration. Since the RRIGS hydration potential is represented in terms of gaussian functions, it might be expected to "soften" the energy landscape. To examine this question, we compare the number of energy evaluations required to accomplish a local minimization of ECEPP/3, E_{TOT}, and ECEPP/3 + MSEED. The test consists of carrying out a single trajectory of the electrostatically driven Monte Carlo (EDMC) algorithm[45] which is a biased search in the space of local minima. Local minimization

TABLE II.
Number of Energy Evaluations per Second for Three Potential Functions.[a]

Polypeptide	Number of amino acids	ECEPP/3	E_{TOT}	ECEPP/3 + MSEED[b]
Enkephalin	5	167	100	21
58–72 fragment of ribonuclease A	15	18	11	7
BPTI	58	1.2	0.7	0.8

[a] Computations were carried out on a single processor of an IBM SP2 computer.
[b] The SRFOPT hydration parameters[37] were used with MSEED.

TABLE III.
Average Number of Energy Evaluations Required / Local Minimization for Three Potential Functions.

Polypeptide	Number of amino acids	ECEPP/3	E_{TOT}	ECEPP/3 + MSEED[a]
Enkephalin	5	62	50	61
58–72 fragment of ribonuclease A	15	163	140	193

[a]The SRFOPT hydration parameters[37] were used with MSEED.

was carried out with the secant unconstrained minimization solver (SUMSL) of Gay.[46] The average numbers of energy evaluations per local minimization for a large number of minimizations are reported for two peptides in Table III. For both peptides, E_{TOT} requires fewer evaluations per minimization, which may indicate that this particular hydration potential "smoothes" the energy landscape somewhat.

Two other methods for including hydration in conformational energy calculations of proteins have been reported recently. Stouten et al.[42] introduced a new measure of solvation exposure which they call the "occupation." This occupation for a particular atom is a maximum occupation minus a sum of the volumes of the other atoms which are scaled by a gaussian function of the interatomic distance. The resulting hydration potential shares the mathematical strengths of the RRIGS approximation presented here. However, in the parameterization of their model, they introduced only a limited number of atom types (six), and did not generate parameters for the backbone atoms. They carried out short stochastic dynamics simulations of BPTI *in vacuo* and with their hydration potential (using the GROMOS force field). Whereas the *in vacuo* simulation diverged from the crystal structure, the simulation using the hydration potential remained closer to the crystal structure.

Kurochkina and Lee[40] used an approximation to the SASA approach in which they include the contributions from only double overlaps. The authors acknowledged that this is not an accurate approximation, but showed that it correlates with the exact calculation of the SASA. Furthermore, by using a single free energy of hydration density parameter, they found that the resulting free energies of hydration correlated with the Miyazawa and Jernigan[47] contact energies for a number of proteins. The limitations of this method are that they treat all atoms as being the same size, and use only a single free energy density parameter.

The advantage of the RRIGS approximation presented in this work is that it is an accurate approximation of a well-established model for hydration. The test of all of these methods will come when the total conformational energy functions can be minimized globally, and the global minimum energy conformation compared to experimental results. The conformational energy function described here (ECEPP/3 + RRIGS hydration) is being incorporated into the current implementation of the DEM, and the next step in this work will be testing both the DEM algorithm and the conformational energy function on peptides and small proteins.

Conclusion

A hydration potential has been constructed based on the hydration shell volume model which incorporates a pairwise gaussian functional form to approximate the exact calculation of the contributions to the exposed volume of the hydration shells of a pair of atoms. It has been shown to be accurate for widely different conformations. Empirical free energy density parameters have been determined from experimental free energies of solvation for small organic molecules. This new potential can be optimized globally by means of the DEM, and is shown to be computationally efficient. It is currently being implemented within the DEM algorithm and will be used in further testing of this method.

Acknowledgments

This work was supported by NIH Grant GM-14312 and NSF Grant DMB90-15815. Support was also received from the National Foundation for Cancer Research and the Association for International Cancer Research. One of the authors (J.D.A.) is supported by a special fellowship from the Leukemia Society of America. All computations reported here were carried out on the IBM SP2

computer at the Cornell National Supercomputing Facility, a resource of the Cornell Center for Theory and Simulation in Science and Engineering which receives major funding from the National Science Foundation, New York State, the IBM Corporation, and members of its Corporate Research Institute. We thank J. Kostrowicki for many helpful discussions, D. R. Ripoll and A. Liwo for assistance in implementing the RRIGS hydration potential in the ECEPP program, K. D. Gibson for providing his program for calculating the volume of a collection of spheres by means of the algorithm described in Ref. 14, and K. D. Gibson and J. Kostrowicki for helpful comments on this article.

References

1. C. B. Anfinsen, *Science*, **181,** 223 (1973).
2. C. B. Anfinsen and H. A. Scheraga, *Adv. Prot. Chem.*, **29,** 205 (1975).
3. J. Kostrowicki and H. A. Scheraga, *J. Phys. Chem.*, **96,** 7442 (1992).
4. R. J. Wawak, K. D. Gibson, A. Liwo, and H. A. Scheraga, *Proc. Nat. Acad. Sci. USA* (in press).
5. F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79,** 2361 (1975).
6. G. Némethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.*, **87,** 1883 (1983).
7. G. Némethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga, *J. Phys. Chem.*, **96,** 6472 (1992).
8. Y. Ding, D. N. Bernardo, K. Krogh-Jesperson, and R. M. Levy, *J. Phys. Chem.*, **99,** 11575 (1995).
9. K. D. Gibson and H. A. Scheraga, *Proc. Nat. Acad. Sci. USA*, **58,** 420 (1967).
10. A. J. Hopfinger, *Macromolecules*, **4,** 731 (1971).
11. Z. I. Hodes, G. Némethy, and H. A. Scheraga, *Biopolymers*, **18,** 1565 (1979).
12. Y. K. Kang, G. Némethy, and H. A. Scheraga, *J. Phys. Chem.*, **91,** 4105, 4109, 4118 (1987).
13. Y. K. Kang, K. D. Gibson, G. Némethy, and H. A. Scheraga, *J. Phys. Chem.*, **92,** 4739 (1988).
14. K. D. Gibson and H. A. Scheraga, *Mol. Phys.*, **62,** 1247 (1987); **64,** 641 (1988).
15. B. Lee and F. M. Richards, *J. Mol. Biol.*, **55,** 379 (1971).
16. A. Shrake and J. A. Rupley, *J. Mol. Biol.*, **79,** 351 (1973).
17. T. J. Richmond and F. M. Richards, *J. Mol. Biol.*, **119,** 537 (1978).
18. W. Kabsch and C. Sander, *Biopolymers*, **22,** 2577 (1983).
19. H. Wang and C. Levinthal, *J. Comput. Chem.*, **12,** 868 (1991).
20. S. M. Le Grand and K. M. Merz, Jr., *J. Comput. Chem.*, **14,** 349 (1993).
21. M. L. Connolly, *J. Appl. Cryst.*, **16,** 548 (1983).
22. T. J. Richmond, *J. Mol. Biol.*, **178,** 63 (1984).
23. W. Hasel, T. F. Hendrickson, and W. C. Still, *Tetrahedron Comp. Meth.*, **1,** 103 (1988).
24. G. Perrot and B. Maigret, *J. Mol. Graph.*, **8,** 141 (1990).
25. L. R. Dodd and D. N. Theodorou, *Mol. Phys.*, **72,** 1313 (1991).
26. L. Wesson and D. Eisenberg, *Prot. Sci.*, **1,** 227 (1992).
27. G. Perrot, B. Cheng, K. D. Gibson, J. Vila, K. A. Palmer, A. Nayeem, B. Maigret, and H. A. Scheraga, *J. Comput. Chem.*, **13,** 1 (1992).
28. B. von Freyberg and W. Braun, *J. Comput. Chem.*, **14,** 510 (1993).
29. D. A. Liotard, G. D. Hawkins, G. C. Lynch, C. J. Cramer, and D. G. Truhlar, *J. Comput. Chem.*, **16,** 422 (1995).
30. F. Eisenhaber and P. Argos, *J. Comput. Chem.*, **14,** 1272 (1993).
31. S. Sridharan, A. Nicholls, and K. A. Sharp, *J. Comput. Chem.*, **16,** 1038 (1995).
32. D. Eisenberg and A. D. McLachlan, *Nature*, **319,** 199 (1986).
33. T. Ooi, M. Oobatake, G. Némethy, and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, **84,** 3086 (1987).
34. D. T. Jones, W. R. Taylor and J. M. Thornton, *Nature*, **358,** 86 (1992).
35. C. A. Schiffer, J. W. Caldwell, P. A. Kollman, and R. M. Stroud, *Mol. Simul.*, **10,** 121 (1993).
36. Y. Wang, H. Zhang, and R. A. Scott, *Prot. Sci.*, **4,** 1402 (1995).
37. J. Vila, R. L. Williams, M. Vásquez, and H. A. Scheraga, *Prot. Struct. Funct. Gen.*, **10,** 199 (1991).
38. R. J. Wawak, K. D. Gibson, and H. A. Scheraga, *J. Math. Chem.*, **15,** 207 (1994).
39. G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, *Chem. Phys. Lett.*, **246,** 122 (1995).
40. N. Kurochkina and B. Lee, *Prot. Eng.*, **8,** 437 (1995).
41. J. A. Grant and B. T. Pickup, *J. Phys. Chem.*, **99,** 3503 (1995).
42. P. F. W. Stouten, C. Frömmel, H. Nakamura, and C. Sander, *Mol. Simul.*, **10,** 97 (1993).
43. M. Gerstein, J. Tsai, and M. Levitt, *J. Mol. Biol.*, **249,** 955 (1995).
44. M. H. Abraham, J. Andonian-Haftvan, G. S. Whiting, A. Leo, and R. S. Taft, *J. Chem. Soc. Perkin Trans.*, **2,** 1777 (1994).
45. D. R. Ripoll and H. A. Scheraga, *Biopolymers*, **27,** 1283 (1988).
46. D. M. Gay, *ACM Trans. Math Software*, **9,** 503 (1983).
47. S. Miyazawa and R. L. Jernigan, *Macromolecules*, **18,** 534 (1985).